

## What is cooking at AGIL?

**Dr Paul VanRaden and chef Dan Null**  
**USDA Animal Genomics and Improvement Lab,**  
**Beltsville, MD 20705**  
**Email: [paul.vanraden@ars.usda.gov](mailto:paul.vanraden@ars.usda.gov)**  
**Web site: <https://aipl.arsusda.gov/>**



Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (1)

AGIL – VanRaden

## Menu for the talk

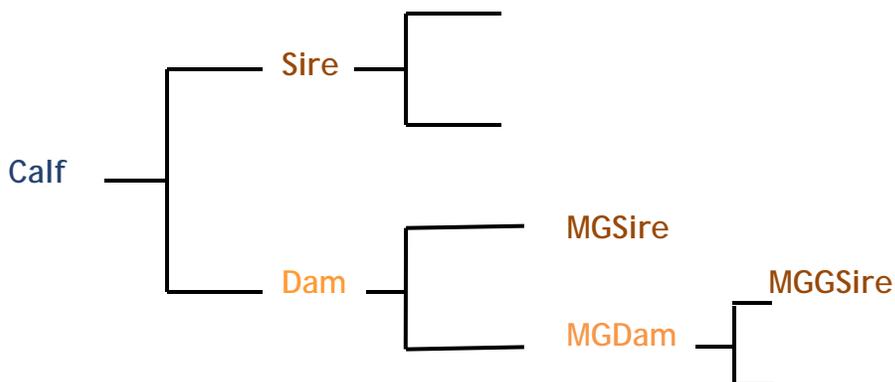
| Course         | Selections  | Price        |
|----------------|---|--------------|
| Appetizer      | Discovering maternal grandsires and maternal great grandsires | Free         |
| Soup and salad | Modeling the accuracy of pedigree data                        | Free         |
| Main course    | New cattle DNA reference map, SNP list, and haplotypes        | Free         |
| Drinks         | Age at first calving as a new trait                           | Free         |
| Dessert        | Feed intake reliabilities                                     | Free         |
| After dinner   | What's cooking at Paul's house?                               | \$2.18 / day |



Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (2)

AGIL – VanRaden

## Appetizer: Discovering pedigrees (parents, grandparents, etc.)



Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (3)

AGIL – VanRaden



## Accuracy of ancestor discovery

- If the true ancestor is genotyped, accuracy of discovery is very high
- If the reported pedigree is wrong, accuracy of detection also high
- With lower density chips, accuracy is slightly reduced
- Methods published and used since 2013:

[https://aipl.arsusda.gov/publish/jds/2013/96\\_1874.pdf](https://aipl.arsusda.gov/publish/jds/2013/96_1874.pdf)

| Ancestor | Correctly discovered |
|----------|----------------------|
| SIRE     | ~100%                |
| MGS      | 97%                  |
| MGGS     | 92%                  |

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (4)

AGIL – VanRaden



## Filling in pedigrees



- >200,000 discovered **MGS** not yet used because no **Dam** ID available
- >200,000 discovered **MGGs** not yet used because no **MGDam** ID available
- Procedures developed to fill the missing IDs to link to ancestors
  - **HO999DAM000000001** for example
  - Similar system used for several years in CAN, ID's not released to public
  - Should evaluations and / or public use the discovered pedigrees?
- First check if **true dam** can be discovered in same herd (20,000 found)
  - Match birth and fresh dates, only 1 dam's pedigree matches calf's

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (5)

AGIL – VanRaden



## Accounting for pedigree accuracy



- Bulls differ in accuracy of daughter pedigree
  - Most bulls have about **90%** of 'daughters' correctly coded
  - Some bulls now have **100%** correct because all are genotyped
- New methods were tested to remove biases from reporting accuracy
  - Genes assumed **90%** from reported sire, **10%** from unknown sire group
  - Biases in Holstein PTAs were smaller than expected
  - Changes in rank from new model were also small
- Project completed by AGIL postdoc Heather Bradford (now at VPI)

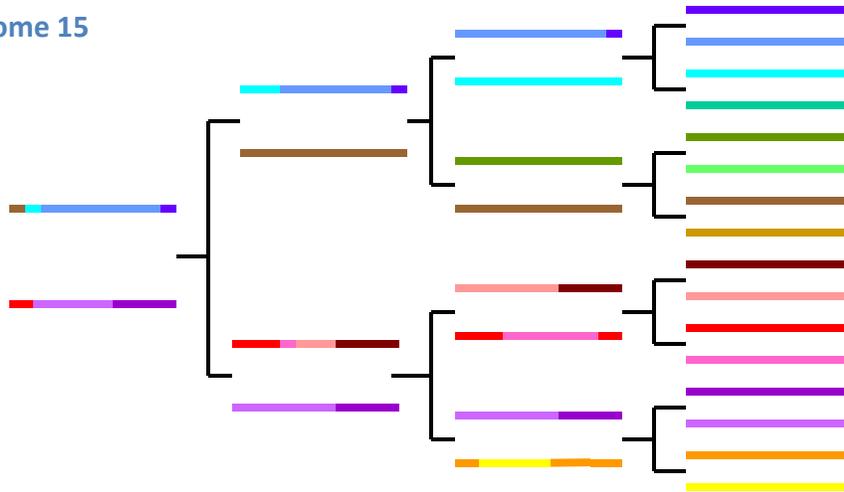
Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (6)

AGIL – VanRaden



## Chromosome Pedigree (O-Style)

chromosome 15



Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (7)

AGIL – VanRaden



## Main course: Cattle DNA reference map



- Why do researchers use the same map?
  - Lets everyone track genetic differences using a common language
  - Shows where genes are, and how the DNA encodes proteins (**annotation**)
- Why switch to a new map?
  - Many sections of the previous map were on wrong chromosome
  - Improve **imputation** of missing SNPs by large-scale rearrangements
  - Improve **annotation** and **alignment** by small-scale refinements
  - 1000 Bull Genomes Project will switch from UMD3 to ARS-UCD

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (8)

AGIL – VanRaden



## Two assemblies: UMD3 and ARS-UCD

- **UMD3**: Zimin et al. **2009** Genome Biology 10:R42
  - University of Maryland and USDA-ARS research
  - Used by most cattle genotyping and sequencing studies since **2009**
  - ‘**Corrected**’ version was used by AGIL after mismapped sections removed
- **ARS-UCD**: Rosen et al. **2018** WCGALP, vol. Molecular Genetics 3, p. 802
  - USDA-ARS-AGIL, U. California-Davis, and several other researchers
  - Used long reads to bridge across repeats, then short reads for polishing
  - Annotation (gene structure) released by NCBI in **May 2018**

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (9)

AGIL – VanRaden



## New cattle genome reference map

- New and old maps both about 2.7 billion bases long
- **ACGTGTCATCGATCCTAGGCTTAGCTACGATCCGATCGATTTCGATCGATCGAATCG**
- Compare previous UMD3 map to new ARS-UCD map
  - Both maps assembled the genome using only **Dominette** DNA
  - Inheritance and consistency of haplotypes across generations
  - Imputation and alignment tests
- Implement the ARS-UCD map during a full release, hopefully December 2018



Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (10)

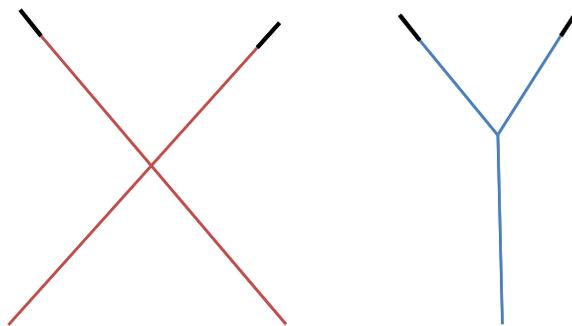
AGIL – VanRaden



## Converting locations to a different map

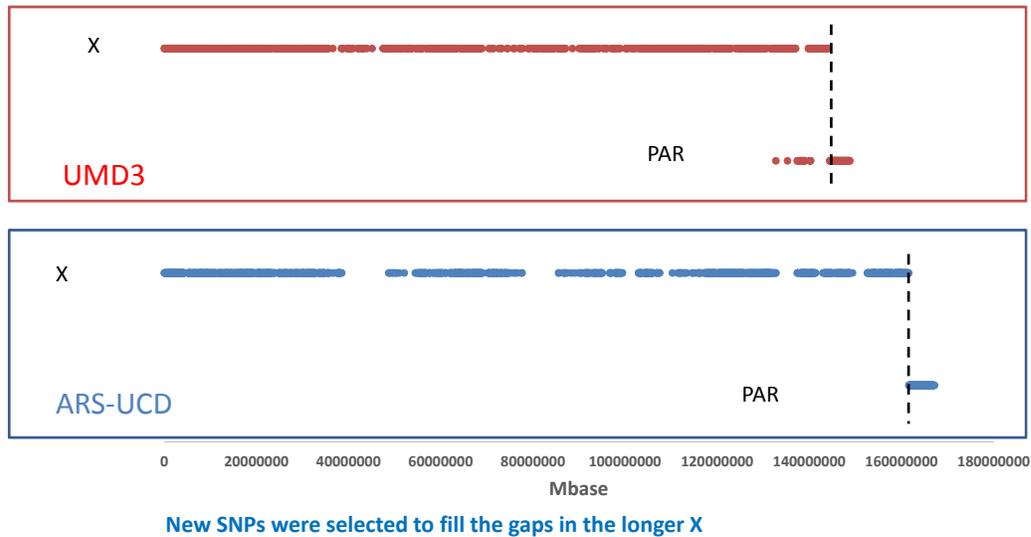
- Convert SNP locations (liftover)
  - Use probe from chip manifest or flanking sequence, align to new map
  - Or simulate paired end reads from old map, align to new map
  - Help from Bob Schnabel, U MO
- Each chromosome may be longer or shorter than previous assembly
  - Many insertions, deletions, inversions, translocations vs. UMD3
  - Autosomes range from **-4.3%** shorter (chr12) to **0.6%** longer (chr26)
  - X chromosome **6.6%** longer than UMD3, with more distinct PAR

## X and Y chromosomes



DNA differs on X and Y,  
but pseudo-autosomal  
region is the same DNA  
on both chromosomes

## Compare pseudo-autosomal region (PAR) of X



Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (13)

AGIL – VanRaden



## Liftover of SNP locations

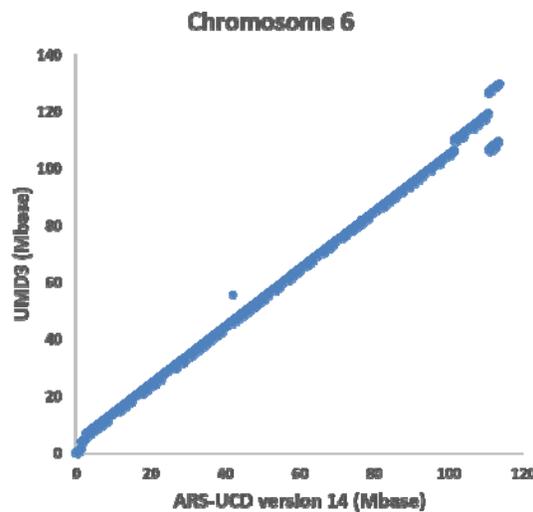
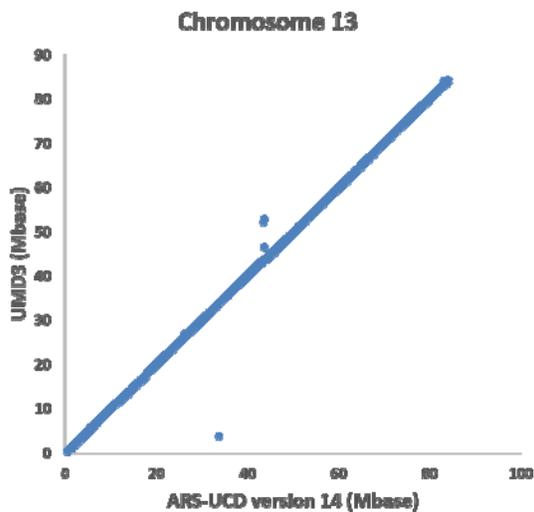
- Probes used for genotyping SNPs on arrays
  - A 50-base sequence extended to the right or left of SNP, such as
  - [C/T]AGTCAGCTCTGTGGCCTGGGCAGGTTCCCAGGATATTCCAGCCAGAC
  - Where was this located on the old map?
  - Where is this located on the new map?
- Difficulties
  - Could have multiple locations in map or other SNPs within the probe
  - Some read errors, or maternal vs. paternal chromosome of Dominette

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (14)

AGIL – VanRaden



## Best and worst chromosome matches (other than X)

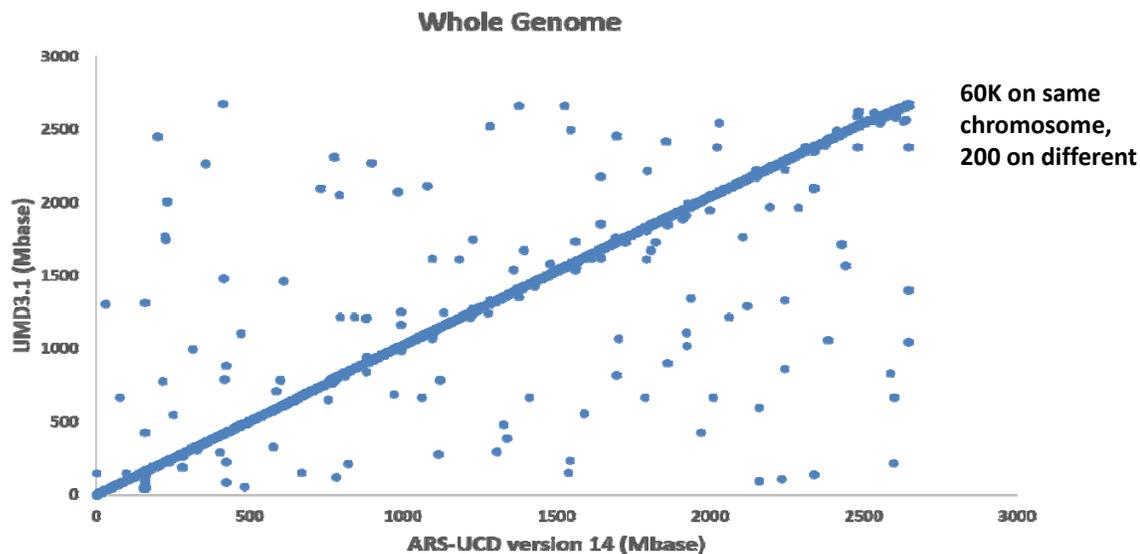


Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (15)

AGIL - VanRaden



## SNPs now located on different chromosomes



Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (16)

AGIL - VanRaden



## Imputation tests



- Locations for the 60,367 usable SNPs were converted to ARS-UCD1
- Genotypes were imputed to 60K from 30 chips of differing density
- Genotypes from 5 breeds were imputed separately:
  - 1,748,453 Holsteins (HO)
  - 215,800 Jerseys (JE)
  - 32,724 Brown Swiss (BS)
  - 4,834 Ayrshires (AY)
  - 3,517 Guernseys (GU)

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (17)

AGIL – VanRaden



## Haplotype noninheritance rate (smaller is better)

| Haplotypes with parent-progeny noninheritance (%) |        |         | Maximum number of haplotypes per segment |        |         |
|---|--------|---------|--|--------|---------|
| Breed   | UMD3   | ARS-UCD | Breed                                    | UMD3   | ARS-UCD |
| HO  | 1.6    | 1.1     | HO                                       | 47,978 | 36,690  |
| JE  | 4.3    | 3.8     | JE                                       | 39,628 | 29,732  |
| BS  | 1.5    | 1.2     | BS                                       | 11,602 | 9,447   |
| AY  | 1.8    | 1.4     | AY                                       | 2,030  | 1,606   |
| GU  | 1.6    | 1.3     | GU                                       | 1,970  | 1,427   |
| <b>SNPs used</b>                                  | 60,671 | 60,474  |  |        |         |

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (18)

AGIL – VanRaden



## Imputation and alignment: Other properties

- Average numbers of distinct haplotypes per segment decreased for all breeds.
- Many previous problem areas no longer have excess numbers of haplotypes, particularly on the **X** chromosome and the pseudoautosomal region of **X**.
- Truly lethal haplotypes were more cleanly separated from false candidate haplotypes.
- To test alignment of sequence data, paired-end reads from a HO bull were aligned to both maps
  - 2.3% more read pairs of Holstein DNA aligned correctly to the new Hereford reference map

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (19)

AGIL – VanRaden



## Implementing new map with more SNPs

- New list of **79,239** instead of **60,671** SNPs
  - SNPs from high density and sequence data with biggest NM\$ effects
  - More gene tests (**DGAT1**, **ABCG2**, **casein**, **leptin**) from recent chips
  - Improved SNP editing system and improved prediction reliability (**~2%**)
- Impute genotypes for all breeds and crossbreds
- Obtain prior allele effect estimates for all traits, breeds, and BBR
- Check published haplotypes to ensure proper inheritance

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (20)

AGIL – VanRaden



## Fertility haplotype research

- Changes to genetic recessives and haplotypes
  - More gene tests incorporated (**HCD, HH4, HH5, HH6, BH2, AH1, AH2**)
  - New Holstein haplotype **HH6** discovered in France (0.4% frequency in USA)
  - **JH2** difficult to track with new map, probably will be discontinued
  - **BH1** fertility effect has greatly decreased, probably will be discontinued
- New map and added gene tests should improve imputation from low density
  - Some chips provide gene tests, many do not

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (21)

AGIL – VanRaden



## Drinks: Age at first calving

- Heifers eat feed but produce no milk until calving
  - Earlier calving is worth **\$2.50** per day
  - Economic emphasis could be **3%** of Net Merit \$
  - Removes some emphasis from heifer conception rate
- Large database (**23 million** records) available for AFC
  - Heritability is **2.7%**
  - PTA standard deviation only about **3** days
  - Reliability of genomic predictions of **66%** for HO

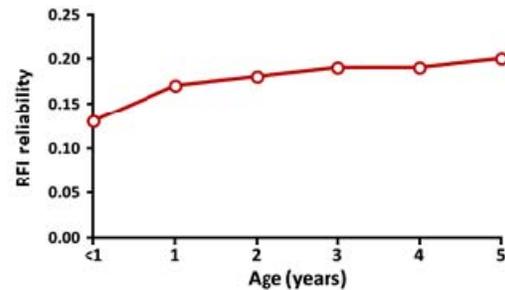
Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (22)

AGIL – VanRaden



## Dessert: Residual feed intake

- Tested potential gains from additional feed intake data
- Estimated young bull reliabilities by exact inversion of single-step genomic equations
- Bulls further from reference cows have lower theoretical REL for RFI
- Measure feed intake on cows more closely related to next generation
- Research of **Bingjie Li** (AGIL postdoc) and **Elif Gunal** (CDCB intern)



Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (23)

AGIL – VanRaden



## What's cooking at Paul's house?



- Least cost ration, every day for 31 years.
- Costs **\$2.18 / day** for food, drink, and snacks
  - Breakfast: 2-day rotation of corn flakes or raisin bran plus whole milk
  - Lunch: every day a bologna sandwich
  - Dinner: 4-day rotation of macaroni, spaghetti, rice, noodles
  - Dessert: every day ice cream
- Invented by me in 1987 at Kohl's food store, Madison, WI
- Guaranteed to prevent both starvation and obesity (100% successful so far)

Council on Dairy Cattle Breeding Industry meeting, Madison, WI, October 2, 2018 (24)

AGIL – VanRaden



## Main points to ruminare on

- Many non-pedigreed herds could get accurate pedigrees
- The **ARS-UCD** sequence map will soon replace the **UMD3** map
  - Better marker locations, genotype imputation, haplotype inheritance, sequence alignment, and gene annotation
  - 30% more SNPs to improve reliabilities and carrier status
- More new traits such as age at first calving could be evaluated
- Feed intake research cows should be related to top new bulls
- Paul's cookbook <http://www.paulvanraden.com/Nutrition.htm>

## Acknowledgments

- USDA-ARS project 1265-31000-101-00, “Improving Genetic Predictions in Dairy Animals Using Phenotypic and Genomic Information” (**AGIL funding**)
- Council on Dairy Cattle Breeding (**CDCB**) and its industry suppliers for data
- Agriculture and Food Research Initiative Competitive Grant #2011-68004-30340 from USDA National Institute of Food and Agriculture (**feed intake funding**)
- Genotyping labs have added many important SNPs to their chips